Signal, Image and Video Processing (2022) 16:29–37
https://doi.org/10.1007/s11760-021-01953-y

ORIGINAL PAPER

Detection of early dangerous state in deep water of indoor swimming pool based on surveillance video

Fan Wang[1] · Yibo Ai[1] · Weidong Zhang[1]

Received: 21 August 2020 / Revised: 13 December 2020 / Accepted: 29 May 2021 / Published online: 14 June 2021

Abstract
This paper presents a method for early detection of dangerous condition in the deep-water zone of swimming pool based on video surveillance. We propose feature extraction, feature expression and assessment criteria, including a method for evaluating normal swimming speed based on the time series of swimmers, a method for assessing an upright state that is not limited by the camera angle, and the rules for assessing dangerous state. We have collected real-life data from the swimming pool and conducted related experiments. Our method can easily and efficiently detect the swimmer who is in danger at an early stage and provide necessary rescue reminders to lifeguards.

Keywords Surveillance video · Dangerous state detection · Swimming pool

1 Introduction

As people's awareness of fitness grows, more and more people choose to swim, and the number of indoor swimming pools is increasing. Although many countries have strict requirements for swimming pools, such as a certain number of lifeguards, and strict standards for inclination and depth, drowning accidents in indoor swimming pools are common. The main causes of accidental drowning in indoor swimming pools are: (1) swimming after drinking or eating after satiety, which causes heart or cardiovascular diseases; (2) insufficient warm-up causes drowning due to seizures; (3) the swimmer exhibits no apparent struggle and stress when drowning, with the result that lifeguards failing to rescue in time.

The primary method used by lifeguards to determine whether drowning has occurred is based on their own rescue experience.

The main characteristics of drowning in humans are: (1) The head is mainly on the water, and the mouth is on a horizontal surface; (2) the head is thrown back, the mouth is open; (3) the eyes are open and dull, and the mouth is half closed; (4) eyes are closed and expressionless; (5) shortness of breath; (6) stand in the water or upright in water; (7) arms extend to the sides or forward.

Since most people cannot call for help and wave their hands when drowning, and there are too many people swimming in the pool, lifeguards may not be able to find swimmers who are in danger in time. When the swimmer's body sinks to the bottom of the pool, it is already at a late stage of drowning and may miss the best time to rescue. Lifeguards cannot keep an eye on every swimmer, but surveillance video can do it. With the development of intelligent surveillance and video surveillance, video surveillance can be used for target detection, tracking, and evaluate complex target behavior. Therefore, it is very important to use video surveillance in the pool to help lifeguards assess the danger of swimmers and reduce the number of accidents.

Indoor swimming pools usually divided into two types: deep water and shallow water. In deep water, swimmers cannot stand in the water, and their legs can only stay for a moment. When stepping, only the swimmer's head is exposed. In shallow water, swimmers can stand directly in the pool, swimmers may not swim, and adults' shoulders can be exposed when they stand up. Due to the large differences between the two, this article mainly discusses the detection of dangerous condition for deep water. The difference between deep and shallow water is shown in Fig. 1.

In this article, we propose a dynamic speed assessment method for assessing whether a swimmer's speed has changed dramatically, and a method for assessing the

✉ Weidong Zhang
zwd@ustb.edu.cn

[1] National Center for Materials Service Safety, University of Science and Technology Beijing, Beijing 100083, China

(a) **(b)**

Fig. 1 **a** In shallow water, swimmers stand right in the pool with their shoulders and back exposed. **b** In deep water, swimmers bend their legs while treading and only their head exposed

swimmer's upright status without considering the camera angle, as well as rules for assessing the dangerous state of swimmers in deep water.

The paper is organized as follows. Section 2 presents the related work. Section 3 introduces our method of assessing the dangerous condition. Sections 4 and 5 present experimental results and conclusions.

2 Related work

2.1 Drowning detection

Video images and other sensors can be used to detect drowning swimmers.

The method of using sensors is generally to carry sensor equipment on the swimmer: (1) Use sonar to detect the swimmer's position and speed, and whether the lungs are filled with water to judge drowning. (2) Use acceleration sensors on the swimmer's wrist and back to detect the swimmer's arm swing frequency, amplitude, swimming speed [1], and whether it is struggling to make drowning judgment. (3) Use photoelectric sensors to detect whether the swimmer's heart rate and blood oxygen [2] are normal. (4) Detect whether the swimmer's head is out of the water, etc. by placing sensors on the head [3], ears [4], etc.

There are three main ways to detect drowning using video surveillance: (1) use underwater cameras to monitor whether the swimmer's body is submerged in water; (2) a fixed camera is used to detecting the location of violent splashes on the water, and a rotary underwater camera is used assessing the position of the splashes by underwater swimmers; (3) drowning characteristics are used to assess drowning only through video surveillance [5–8].

Most indoor swimming pools are not equipped with underwater cameras, and underwater video surveillance method for drowning detection cannot be quickly promoted. In a study of drowning detection using original water surveillance video, a team from Singapore has done a series of very significant works. In an early study in 2002 [5], they used Gaussian Mixture Model to simulate swimmer and background modeling for swimmer detection, and segmented the swimmer's body shape. They used Kalman filter for swimmer tracking, and extracted three swimmer features (moving speed, size variation, elongation measure) and three drowning assessment rules for drowning detection. In order to efficiently detect and segment swimmers, the team proposed a block-based background modeling and a hysteresis threshold method in 2003 [7], and extracted five swimmer descriptors (speed, posture, submersion index, activity index, splash index) and built a functional link network. They proposed the first live visual surveillance system for early drowning detection in a pool (DEWS) in 2008 [8], and proposed a set of methods in the areas of background subtraction, denoising, data fusion and blob splitting, and a module comprising data fusion and hidden Markov modeling to study unique features of various swimming behaviors, in particular, these early drowning cases. They used the drowning experiment from the laboratory in the real world.

The methods used in previous studies do not adequately achieve our goals, as they suffer from one or more of the following problems:

(1) In the sensor drowning detection method, swimmers must carry sensor devices, which may not be well implemented in large public swimming pools, and sensor charging must also be considered.
(2) In the underwater video surveillance method, the swimmer's body sinks to the bottom only in the late stage of drowning. Hence, the method detecting if a swimmer's body is submerged in the water can easily miss the best rescue time.
(3) The main problems with video surveillance methods are: In terms of swimmer detection, the accuracy of the traditional background and swimmer modeling methods is still not very accurate, and the error increases when the background is very complex or changes abruptly. In terms of extracting drowning features, the ratio of the displacement of the swimmer's center of mass to time is used to approximate the swimmer's speed. There is a big difference between the near camera end and the far camera. The long and short axis of the outer ellipse of the swimmer's shape is used to determine whether the swimmer is upright. This method does not work if the optical axis of the camera is not perpendicular to the swimmer's swimming direction.

In summary, the main purpose of this article is to use a higher precision method to detect the position of swimmers. The characteristics of swimmers are extracted by fully combining the temporal characteristics of the video. To use a better upright feature to meet different angle cameras of indoor swimming pools.

2.2 Object detection

With the support of the GPU and other hardware, deep-learning algorithms are widely used in the field of computer vision, such as classification problems, target detection problems, and segmentation problems. Compared with traditional object detection methods based on machine learning algorithms, it can learn functions automatically at multiple levels.

Basically, there are two types of object detection methods.

The first one is a two-step method. The basic idea is to extract candidate regions from the image, and then process the candidate regions to determine the detection target. The most typical two-step method is the R-CNN series: In 2014, Ross Girshick [9] proposed the R-CNN method. First, the candidate area method is used to create areas of interest, and then those areas are converted into fixed-size images and sent to convolution. In the neural network, multiple fully connected layers are tracked to classify objects and refine the bounding box. Since candidate area extracted in R-CNN have overlapping parts, which affects the speed of training. Ross Girshick [10] proposed Fast-RCNN to directly extract the features of the entire image. In order to increase the running speed, Ren [11] proposed Faster-RCNN, which replaced the candidate area method with an internal deep network.

The second type of object detection methods is a one-step method that can predict multiple locations and categories of objects at once, and can achieve end-to-end object detection and recognition. The biggest advantage of this method is its speed. Typical one-step methods include the YOLO [12–14] series proposed by Joseph Redmon et al., and the SSD [15] proposed by Liu Wei et al. YOLOv3 [14] of the YOLO series used the residual network Darknet-53 network of the residual network structure to extract image characteristics, forming a deeper network layer, and used 3 different scaled maps for object detection. Feature maps of different scales are suitable. Targets of different sizes improve the detection effect of small objects.

3 Method

All surveillance videos used in this article were captured from two surveillance cameras in the southwest corner of a swimming pool in Beijing. These two cameras are mainly responsible for monitoring the three swimming lanes in the deep-water area. The swimming lanes are 2.4 m deep and 2.5 m wide. Most of the swimmers in the video swim autonomously, and some of them swim in typical experimental positions.

3.1 Head detection based on YOLOv3

The surveillance video was recorded at a rate 25 frames per second, in total 5033 images were intercepted, with a resolution of 640*480. According to the three lanes monitored by the camera, each lane used 1511 images for training, and the remaining 500 images were reserved for testing. The rectangular box was used to label the head of the swimmer in the picture, and the YOLOv3 deep-learning network was used to train 20,000 times to obtain a head detection model. The detection effect is shown in Fig. 2.

3.2 Normal speed assessment based on head bounding box

Due to the angle of the camera and the nature of the image, the distance between the far-end camera and the near-end camera in the image is very different from the actual distance. Therefore, it is not possible to directly judge the actual displacement of the swimmer by the change in the displacement of the swimmer's head in the image. Video surveillance captures four different states of freestyle, breaststroke, backstroke, and tread water (each about 16 s, about 400 frames). Using the head detection model, the bounding box of the swimmer's head is obtained, and the change in the area of the bounding box is shown in Fig. 3.

In the first three conditions, the area of the bounding box decreases as the swimmer moves away from the camera, while in treading, the area of the bounding box continues fluctuate within a certain range. The position of the swimmer's head in the video is shown in Fig. 4.

When the number of selected frames is between 25 and 50, the area of two adjacent bounding boxes overlaps less. Figure 5 shows the curve of Intersection-over-Union (IoU) of the two adjacent frames where the number of sampling frames is 35.

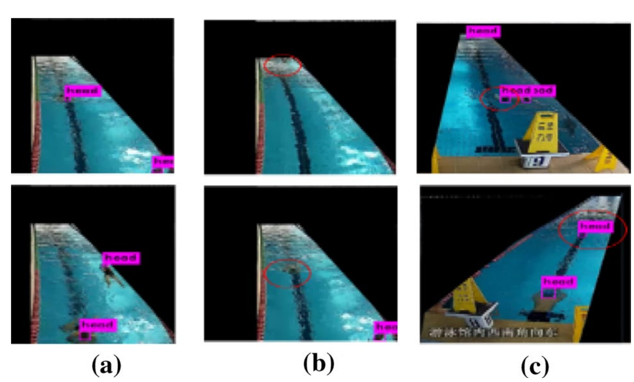

Fig. 2 **a** Image of the head detection result. **b** Lack of head detection. **c** False head detection

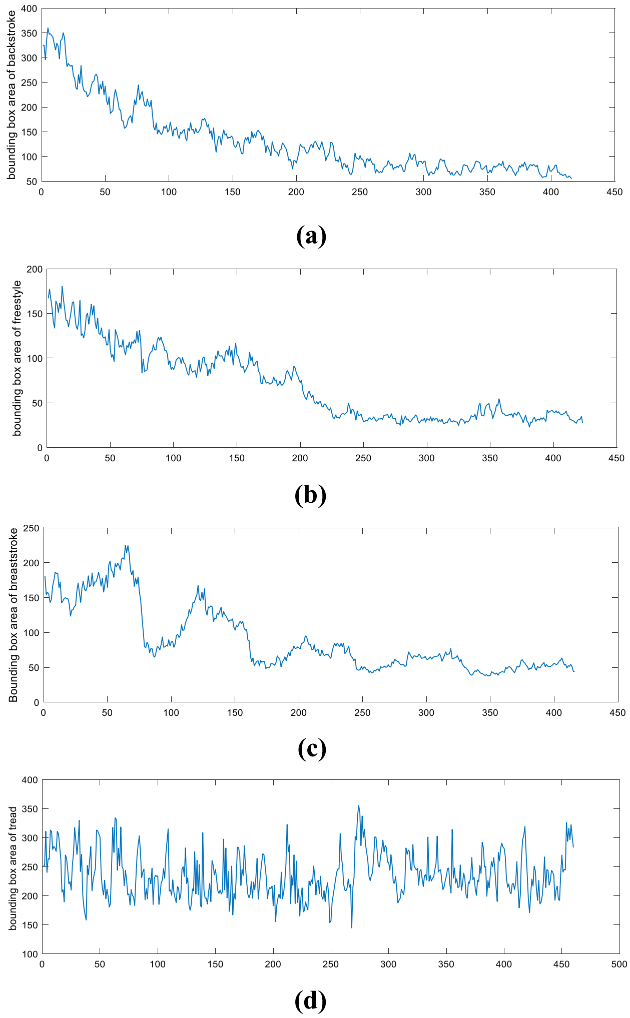

Fig. 3 **a**, **b**, **c**, **d** Head bounding box area for freestyle, breaststroke, backstroke, and tread

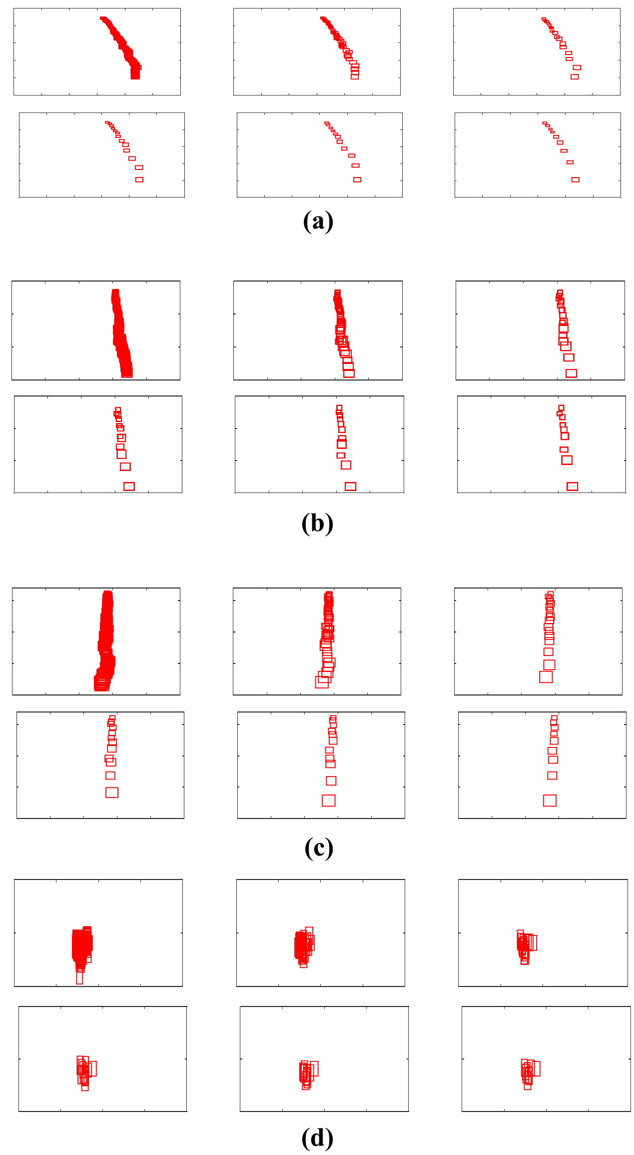

Fig. 4 **a**, **b**, **c**, **d** is the position of the head bounding box in four states, and the sampling interval is 1 frame, 10 frames, 25 frames, 35 frames, 45 frames, and 50 frames. The video is 25 frames per second

The IoU in the swimming state periodically changes in a small range. Since the swimmer practically does not move when treading water, the IoU is large (greater than 0.6). According to the analysis of the swimmer's head bounding box, this paper used the overlapping area of the adjacent swimmer head bounding box to judge whether the swimmer is swimming normally.

Rule to obtain initial sample number t: According to the first m frames of the video, the initial sample frame number t is determined and then update continuously.

The stroke rate for ordinary swimmers is about 0.9–1.7 s each time, the swimming distance is about 2 m each time, m takes 2–4 stroke cycles. This video is 25 frames per second. That is, m takes 100 frames in this experiment.

Find the first frame f_{t_0} that does not overlap the bounding box of frame f_0, then the sample frame number is t_0.

The sample number t of the frame f_n can be calculated as:

$$t = n - \arg\min_i S_{n,i} \tag{1}$$

where $n > 0$ and $i = n - t - 1, n - t, n - t + 1$.

$S_{n,i}$ is the IoU of frame f_n and frame.

Normal speed assessment: Find if the IoU is within the predetermined normal range.

According to the sample frame number t, we can determine if the IoU of the frame f_n and the frame f_{n+t} are within the preset normal range, if yes, then it is in a normal swimming state, and the sample frame number t update as:

$$t = (n + t) - \arg\min_i S_{n,i} \tag{2}$$

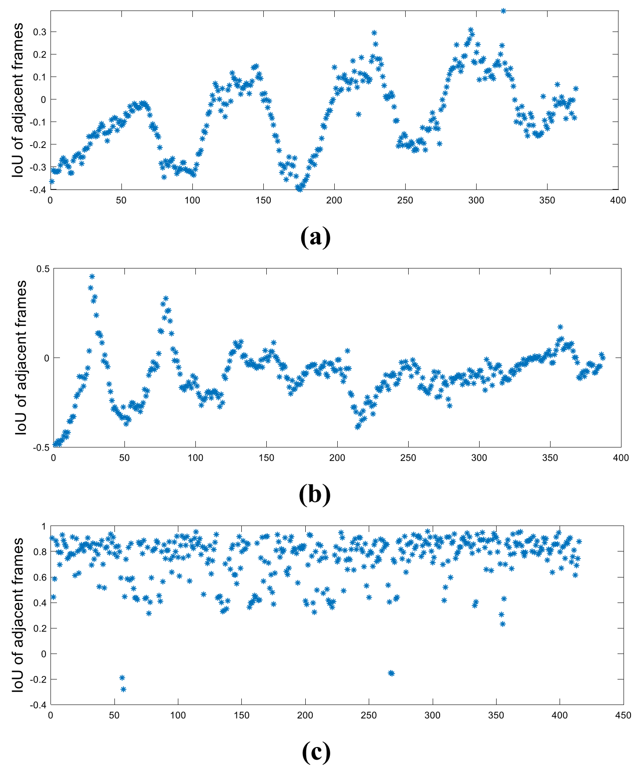

Fig. 5 **a**, **b**, **c** is the IoU curve in the freestyle, breaststroke and treading state

where $i = n - 1, n, n + 1$.

3.3 Assessment of upright dangerous state based on head-to-body ratio

Assessment of upright dangerous state: If swimmer is drowning in deep water, most of his body is upright. To distinguish between treading state and upright dangerous state, combined with video viewing angle in the conventional pool, the tread condition and the upright dangerous state are assessed according to the head-to-body ratio.

The head-to-body ratio of a normal person is about 6.5–7.5. Due to the angle of the camera and refraction, the head-to-body ratio in water is slightly different. However, in the treading state, due to the curled-up state, the head-to-body ratio of the swimmer will be significantly different Fig. 6 compares the head-to-body ratio between the bent legs versus upright position. Since the full straightened state of the legs cannot be achieved with normal treading, the upright state of the legs in the picture is the experimentally simulated state.

Rule to obtain head-to-body ratio: Select the outline of human shape from the outline of the edge.

Firstly, the histogram of the bounding box of the swimmer's head is computed. Then, the rectangle in eight directions is determined (Fig. 7b) according to the center position of the head. We calculate the similarity between the histogram of the eight directions and the head, select the direction with the most similarity, and expand it to the area of interest shown in Fig. 7c. Finally, the edges of the area of interest are extracted, and the outer rectangle of the human body is got and shown in Fig. 7d.

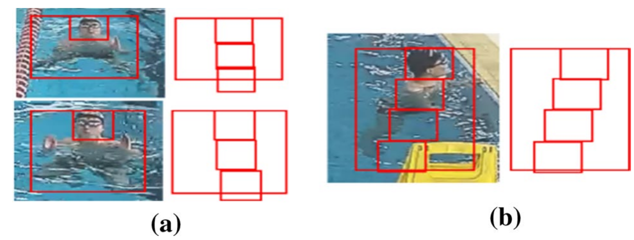

Fig. 6 **a** is the bent legs state when treading water. **b** is the upright state

3.4 Swimmer in danger assessment in deep water

If no head is found within a certain period of time, an alarm is issued. If there is no bounding box that does not overlap with the first bounding box of the head, that means the target is not swimming. Then, we can evaluate if the swimmer is in an upright treading state. If the swimmer is in an upright dangerous state, an alarm will sound. If the number of sampling frames is specified, the normal swimming speed will be evaluated. If it is abnormal, it means that the target's speed has changed. Once the threshold is exceeded, the upright dangerous state is assessed. If it is in an upright dangerous state, an alarm will sound (Fig. 8).

4 Experimental results

We have collected a few videos (each video is about 16 s, about 400 frames) of four different states of freestyle, breaststroke, backstroke and treading.

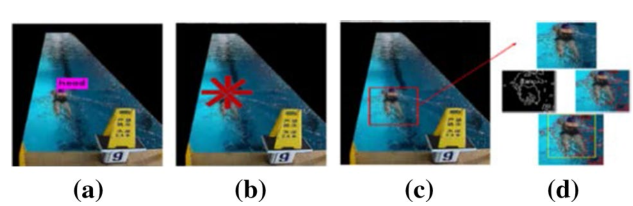

Fig. 7 **a** head position **b** eight directions **c** area of interest **d** the outer rectangle of the human body

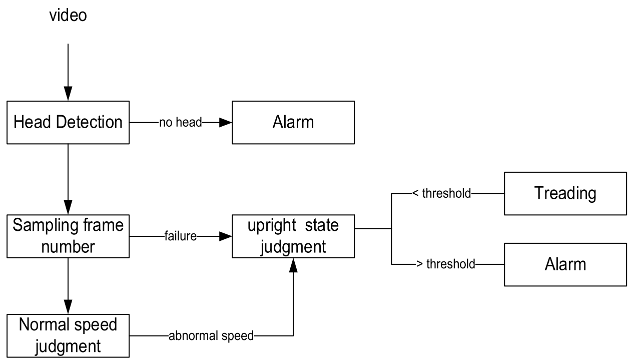

Fig. 8 Swimmer in danger Assessment

4.1 Sample number and IoU

We collected videos of four states: freestyle, breaststroke, backstroke, and treading, and then determine the number of sample frames for the first 100 frames. Some experimental results are shown in Table 1.

For data with numbers 1 and 6, the update of the sample frame number is shown in Fig. 9.

Figure 9a shows a process for updating the sample frame number of data 1. In the 11th frame, there is bounding box that does not overlap with the first frame, that is, the initial sample number t_0 is 11, and then it is updated sequentially, in 100th frames the sample number t_{100} is 11. The swimmer in this data moved almost at a constant speed. Figure 9b shows the IoU changes during the update, in the range (− 0.049, 0.042). Figure 9c shows a process for updating the sample frame number of data 6. In the 51th frame we see the bounding box that does not overlap with the first frame, that is, the initial sampling number t_0 is 51, and then it is updated sequentially, in 100th frames the sampling number t_{100} is 85. The swimmer in this data accelerated. Figure 9d shows the IoU changes during the update process, in the range (− 0.237, 0.015).

4.2 Normal speed assessment

Figure 10a shows the process of updating the sample frame number of data 1 at normal speed assessment process. The initial sample number t_0 is 11, and then it is updated sequentially. Figure 10b shows the IoU changes during the normal speed assessment process, in the range (− 0.1, 0.6). Figure 10c shows the process of updating the sample frame number of data 6 at normal speed assessment process. The initial sample number t_0 is 43, and then it is updated sequentially. Figure 10d shows the IoU changes during the normal speed assessment process, in the range (− 0.3, 0.25).

4.3 Head-to-body ratio

We calculated the similarity between the eight-direction histogram and the head and chose the direction with most similarity. Figure 11 shows the information about four random directions. The red curve is the histogram distribution of the head's bounding box, the blue curve is the histogram distribution in that direction. Their correlations are 0.63218, 0.06036, 0.13431, 0.02841.

Table 1 Sample number and IoU

	1 (Freestyle)	2 (Freestyle)
t_0	11	46
Initial sample number	11	85
IoU	(− 0.049, 0.042)	(− 0.187, 0.027)
	3 (Backstroke)	4 (Backstroke)
t_0	51	52
Initial sample number	89	83
IoU	(− 0.237, 0.015)	(− 0.251, 0.052)
	5 (Backstroke)	6 (Breaststroke)
t_0	30	13
Initial sample number	77	43
IoU	(− 0.025, 0.2)	(− 0.073 , 0.066)
	7(Breaststroke)	8 (Treading)
t_0	33	–
Initial sample number	64	–
IoU	(− 0.037, 0.019)	(0.245, 0.810)

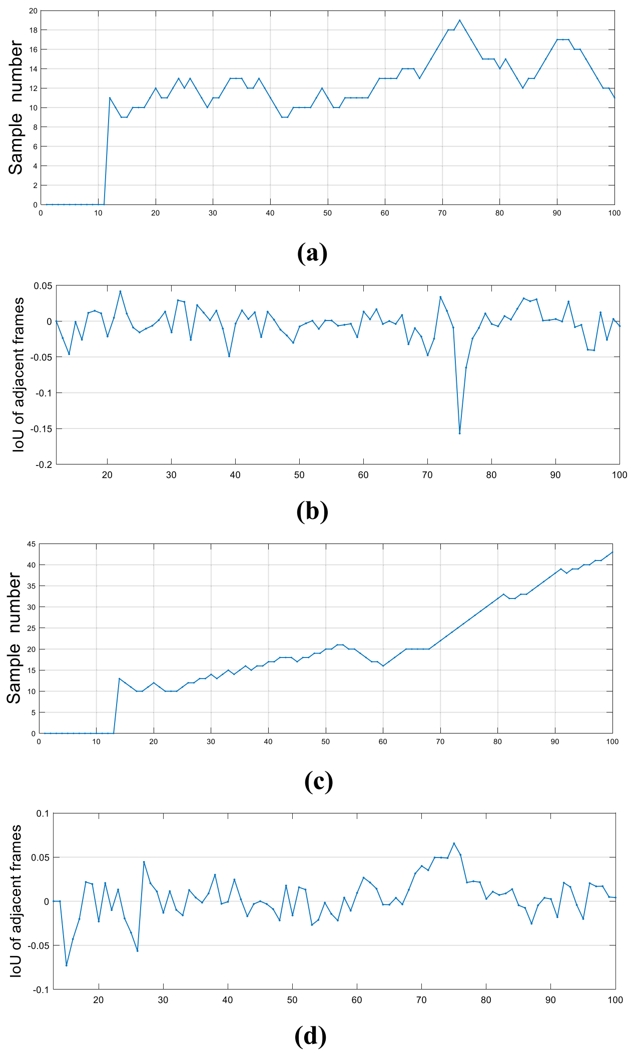

Fig. 9 **a** Process of updating the sample frame number of data 1 (Freestyle). **b** IoU of data 1(Freestyle). **c** The process of updating the sample frame number 6 (Breaststroke). **d** IoU of data 6 (Breaststroke)

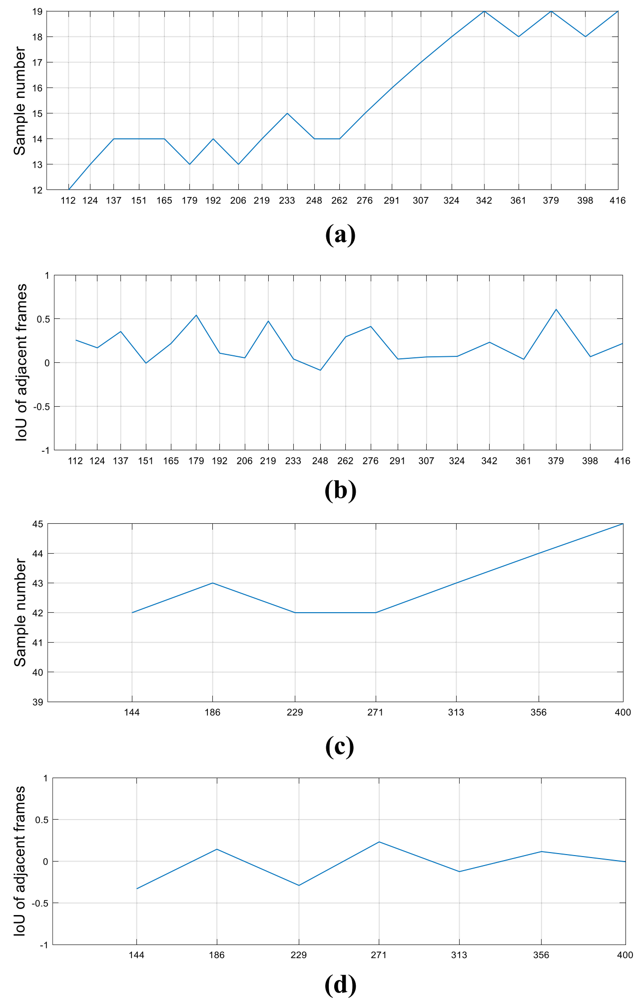

Fig. 10 **a** Process of updating the sample frame number of data 1 (Freestyle). **b** IoU of data 1 (Freestyle) at normal speed of assessment process. **c** The process of updating sample frame number of data 6 (Breaststroke). **d** IoU of data 6 (Breaststroke) at the normal rate of the assessment process

In the case of Fig. 11, the situation of Fig. 11a is the most suitable direction, therefore the operation of expanding the region of interest in this direction was chosen.

We chose the vertical direction with most similarity to expand the area of interest, extracted the edge outline in the area of interest, and found the swimmer's bounding box. The image processing steps are shown in Fig. 12.

We randomly selected images for head-to-body ratio experiments. The head-to-body ratio when treading water is approximately 1:3, and the head-to-body ratio in a normal straightened swimming leg is between 1:6 and 1:9. Some results are shown in Fig. 13.

5 Conclusion

We use the time sequence information of the swimmer's head bounding box to determine if the swimmer is swimming at normal speed according to the overlapping area of the adjacent bounding box, which solves the large errors problem of in the method of directly using the object displacement time ratio. The method of assessment the upright state by using the head-to-body ratio solves the problem that the long and short axis of the outer ellipse of the human body do not work when the optical axis of the camera is not perpendicular to the swimming direction of the swimmer.

Based on the characteristics of drowning and the characteristics of the deep-water area of pool, we propose the method for identifying signs and assessing early dangerous state of swimmers in deep-water zone. Our method makes it

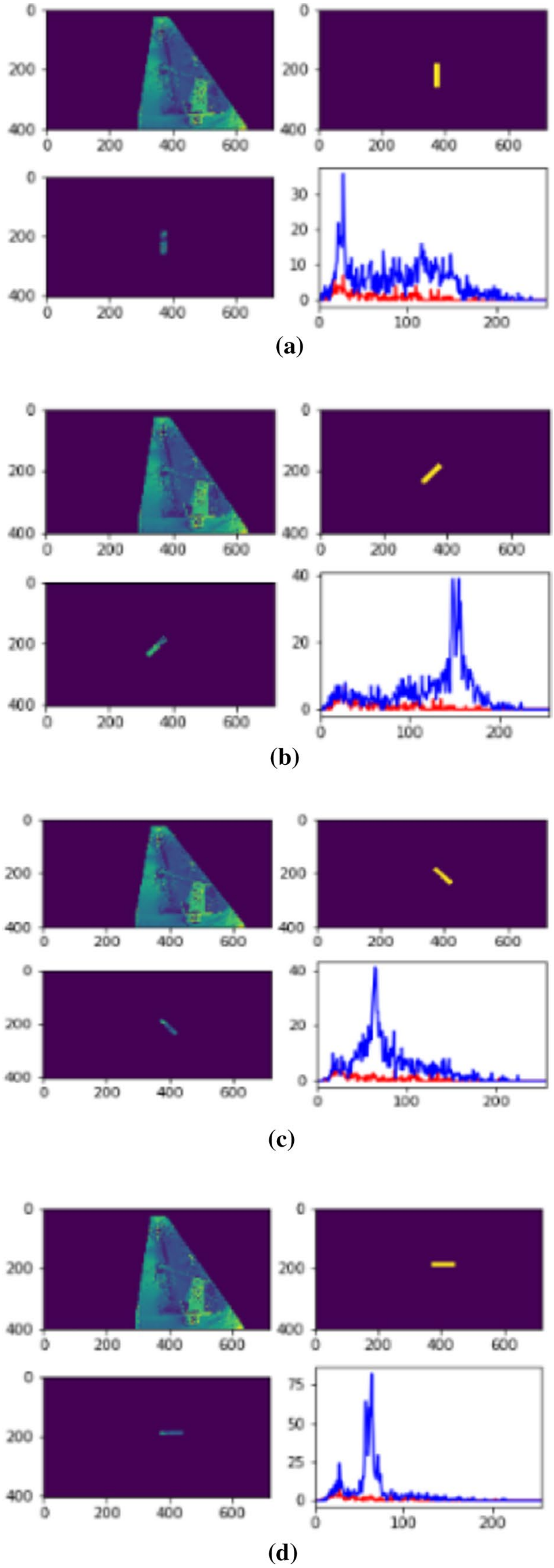

Fig. 11 **a**, **b**, **c**, **d**—histogram information in four directions

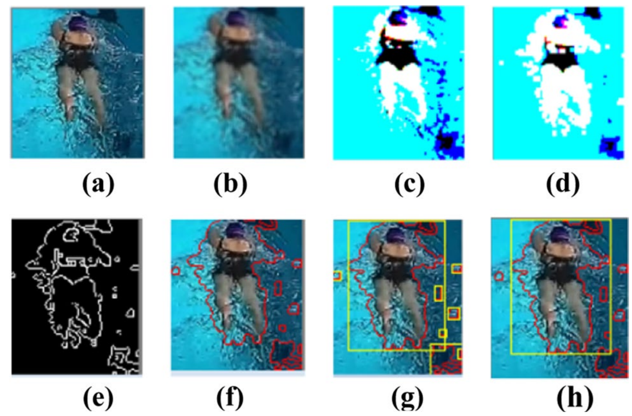

Fig. 12 Image processing steps. **a** original image **b** Gaussian filtering **d** dilate **g** circumscribed rectangle **h** Maximally circumscribed rectangle

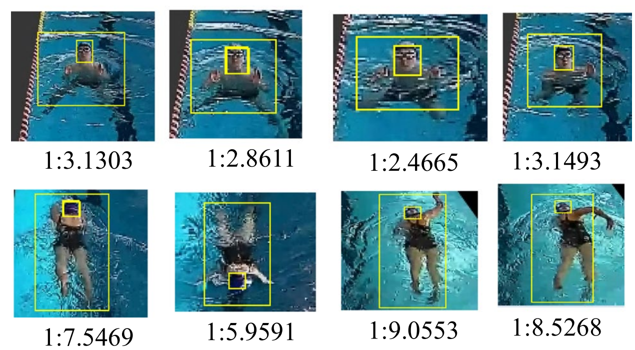

Fig. 13 Head-to-body ratio

easy and effective to screen out swimmers who are in danger at an early stage and provide lifeguards with the necessary reminders for help.

References

1. Delgado-Gonzalo, R., et al.: Real-time monitoring of swimming performance. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4743–4746, (2016) https://doi.org/10.1109/EMBC.2016.7591787
2. Kulkarni, A., Lakhani, K., Lokhande S.: A sensor based low cost drowning detection system for human life safety. In: 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 301–306, (2016) https://doi.org/10.1109/ICRITO.2016.7784970
3. Kharrat, M., Wakuda, Y., Koshizuka, N., Sakamura K.: Near drowning pattern detection using neural network and pressure information measured at swimmer's head level. In: Proceedings of the Seventh ACM International Conference on Underwater Networks and Systems - WUWNet '12, Los Angeles, California, p. 1, (2012) https://doi.org/10.1145/2398936.2398994

4. Hariharan, B., Das, R.N., Arjun, S.: WiEyeTNB: a wireless sensor based drowning detection system for enhanced parental care. Wirel. Internet **98**, 511–520 (2012). https://doi.org/10.1007/978-3-642-30493-4_49
5. Lu, W., Tan, Y-P.: A camera-based system for early detection of drowning incidents. In: Proceedings. International Conference on Image Processing, Toronto, Ont., Canada, vol. 1, p. III-445–III-448, (2003) https://doi.org/10.1109/ICIP.2002.1039001
6. Lu, W., Tan, Y-P.: Swimmer motion analysis with application to drowning detection. In: 2002 IEEE International Symposium on Circuits and Systems. Proceedings (Cat. No.02CH37353), vol. **2**, p. II–II, (2002) https://doi.org/10.1109/ISCAS.2002.1011439
7. Eng, H.L., Toh, K., Wang, J., Yau, W.Y.: An automatic drowning detection surveillance system for challenging outdoor pool environments. In: Proceedings Ninth IEEE International Conference on Computer Vision, vol.1, pp. 532–539, (2003) https://doi.org/10.1109/ICCV.2003.1238393
8. Eng, H., Toh, K., Yau, W., Wang, J.: DEWS: a live visual surveillance system for early drowning detection at pool. IEEE Trans. Circuits Syst. Video Technol. **18**(2), 196–210 (2008). https://doi.org/10.1109/TCSVT.2007.913960
9. Liu, Y., Cheng, M.-M., Hu, X., Wang, K., Bai, X.: Richer convolutional features for edge detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, pp. 5872–5881, (2017). https://doi.org/10.1109/CVPR.2017.622
10. Girshick, R.: Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448, (2015). https://doi.org/10.1109/ICCV.2015.169
11. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017). https://doi.org/10.1109/TPAMI.2016.2577031
12. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016). https://doi.org/10.1109/CVPR.2016.91
13. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, pp. 6517–6525, (2017) https://doi.org/10.1109/CVPR.2017.690.
14. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. *arXiv e-prints*, (2018)
15. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016, vol. 9905, pp. 21–37. Springer International Publishing, Cham (2016)